

# Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care

Isobel M Cameron, John R Crawford, Kenneth Lawton and Ian C Reid

## ABSTRACT

### Background

The 2004 National Institute for Health and Clinical Excellence (NICE) guidelines highlight the importance of assessing severity of depression in primary care.

### Aim

To assess the psychometric properties of the Patient Health Questionnaire (PHQ-9) and the depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for measuring depression severity in primary care.

### Design of study

Psychometric assessment.

### Setting

Thirty-two general practices in Grampian, Scotland.

### Method

Consecutive patients referred to a primary care mental health worker completed the PHQ-9 and HADS at baseline ( $n = 1063$ ) and at the end of treatment ( $n = 544$ ). Data were analysed to assess reliability, robustness of factor structure, convergent/discriminant validity, convergence of severity banding, and responsiveness to change.

### Results

Both scales demonstrated high internal consistency at baseline and end of treatment (PHQ-9  $\alpha = 0.83$  and  $0.92$ ; HADS-D  $\alpha = 0.84$  and  $0.89$ ). One factor emerged each for the PHQ-9 (explaining 42% of variance) and HADS-D (explaining 52% of variance). Both scales converged more with each other than with the HADS anxiety (HADS-A) subscale at baseline ( $P < 0.001$ ) and at end of treatment ( $P = 0.01$ ). Responsiveness to change was similar: effect size for PHQ-9 =  $0.99$  and for the HADS-D =  $1$ . The HADS-D and PHQ-9 differed significantly in categorising severity of depression, with the PHQ-9 categorising a greater proportion of patients with moderate/severe depression ( $P < 0.001$ ).

### Conclusion

The HADS-D and PHQ-9 demonstrated reliability, convergent/discriminant validity, and responsiveness to change. However, they differed considerably in how they categorised severity. Given that treatment decisions are made on the basis of severity, further work is needed to assess the validity of the scales' severity cut-off bands.

### Keywords

depression; measurement; primary care; severity.

## INTRODUCTION

In 2004 the National Institute for Health and Clinical Excellence (NICE) guidelines on the management of depression in primary and secondary care emphasised the importance of measuring depression severity to target the condition with an appropriate intervention.<sup>1</sup> This runs in accordance with the stepped-care approach of managing depression, which consists of five steps, beginning with the recognition of depression in primary care by a GP or practice nurse. Following this, different interventions are advocated according to severity.

NICE guidelines recommend the use of the International Classification of Diseases (ICD-10) criteria for diagnosing and assessing severity of depression. This method involves a symptom count which then falls within progressive categories (mild, moderate, and severe [with or without psychotic symptoms]), corresponding to increased numbers of symptoms identified.<sup>2</sup> While advocating this method, NICE also acknowledges that: 'it is doubtful whether severity can realistically be captured in a single symptom count', and that previous history, family history, associated disability, and availability of social support should also be considered.<sup>1</sup>

The new general medical services' Quality and Outcomes Framework provides incentives to practices for making an assessment of the severity of depression at the outset of a new diagnosis of

*IM Cameron, MA Hons psychology, research fellow; IC Reid, BMedBiol, PhD, FRCPsych, professor, Department of Mental Health; JR Crawford, BSc(Hons) MSc, PhD, FBPSS, CClinPsychol, professor, School of Psychology; K Lawton, FRCGP, Department of General Practice and Primary Care, University of Aberdeen, Aberdeen.*

### Address for correspondence

Ms Isobel M Cameron, Department of Mental Health, University of Aberdeen, Royal Cornhill Hospital, Aberdeen, AB25 2ZH. E-Mail: i.m.cameron@abdn.ac.uk

**Submitted:** 31 March 2007; **Editor's response:** 25 May 2007; **final acceptance:** 4 July 2007.

©British Journal of General Practice 2008; 58: 32–36.

DOI: 10.3399/bjgp08X263794

depression.<sup>3</sup> This rationale ensures a discussion can take place with patients of the relevant treatment options, and provides a baseline from which to monitor progress. Practices are required to use a validated assessment tool for this purpose. Those endorsed are: the Patient Health Questionnaire (PHQ-9),<sup>4</sup> the Hospital Anxiety and Depression Scale (HADS),<sup>5</sup> and the Beck Depression Inventory, second edition (BDI-II).<sup>6</sup> Practices are advised to choose one of these three measures.

While research has been conducted to assess the comparative validity and accuracy of questionnaires at detecting depression,<sup>7,8</sup> the relative validity of scales at categorising severity has not been adequately assessed. With an absence of objective psychometric comparisons between measures, GPs may find it difficult to make an informed choice of measure.

As part of an audit of the Scottish Executive's 'Doing Well by People with Depression' programme, the authors of the current study examined the psychometric properties of the PHQ-9 and the HADS on the same sample of patients. Within the service audit, both measures were completed by patients referred to a primary care mental health worker/therapist (mental health worker) based in primary care settings in Grampian, UK. As a result, the relative reliability, validity, and responsiveness to change of the PHQ-9 and HADS depression subscale (HADS-D) can be assessed.

## METHOD

### Participants

A consecutive sample of adults referred by GPs to mental health workers based in 32 general practices across Grampian participated. Inclusion criteria were: adults with a mild to moderate mental health problem who GPs considered might be interested in, and able to concentrate on, a self-help approach. Exclusion criteria applied to persons: under 16 years of age; with severe or complex mental health problems (for example, psychosis, obsessive-compulsive disorder, and comorbid personality disorder); with a history of violent or threatening behaviour; admitting to suicidal ideation or recent/recurrent self-harm; who were currently misusing drugs/alcohol; or who had previously had more than one referral to clinical psychology.

Data collection was performed prospectively as part of an audit of the mental health workers' service. Unique identifier numbers were allocated to each patient by the mental health worker. The university team did not have access to information that identified individuals.

### Measures

As part of a service audit, primary care patients referred

## How this fits in

Gauging the severity of depression is an imperative in primary care if evidence-based interventions are to be offered. The Quality and Outcomes Framework of the new general medical services contract provides incentives for using one of the following depression severity assessment tools: the PHQ-9, HADS, or BDI-II. Comparison of the PHQ-9 and the HADS in this study indicates that both demonstrate acceptable reliability, convergent and discriminant validity, and responsiveness to change, but that they differ considerably in how they categorise severity. The relationship between measurement of severity using depression assessment tools and ICD-10 severity criteria remains unknown and requires further investigation.

to a mental health worker were asked to complete a questionnaire at baseline, which included the HADS and demographic questions. The mental health worker then conducted the PHQ-9 interview schedule at the first appointment. At the end of treatment, patients completed a further questionnaire which included the HADS and PHQ-9 self-complete version.

The PHQ-9 consists of nine questions designed to correspond to the nine diagnostic criteria for major depressive disorder covered in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).<sup>9</sup> Items are rated from 0 to 3 according to increased frequency of experiencing difficulties in each area covered. Scores are summed and can range from 0 to 27. The score can then be interpreted as indicating either no depression, minimal, mild, moderate, moderately severe, or severe depression.

The HADS consists of 14 items each rated from 0 to 3 according to severity of difficulty experienced. Eight items require reversed scoring, after which depression (HADS-D) and anxiety (HADS-A) subscale totals can be summed. Each subscale score can range from 0 to 21. The scores can then be interpreted as indicating mild, moderate, or severe difficulty.

**Table 1. Sample characteristics.**

Characteristic	n (%) <sup>a</sup>
Mean age, years (SD)	37.7 (13.2)
Males	280 (26)
Main activity	
Employed or self-employed	590 (59)
Retired	63 (6)
Housework	141 (14)
Student	73 (7)
Seeking work	61 (6)
Other	78 (8)
Educated beyond minimum school age	572 (57)
Educated to degree level	345 (35)

<sup>a</sup>Maximum n = 1063.

### Statistical methods

Internal consistency of both the PHQ-9 and the HADS-D was examined using Cronbach's  $\alpha$  and item-total correlations. Principal components analysis was used to assess the homogeneity of the scales: separate principal components analyses were performed for each scale at both time points; coefficients of congruence were used to compare factor loadings across the two time points.<sup>10</sup> Correlations of the HADS-D and the PHQ-9 with the HADS-A were calculated to assess whether the PHQ-9 and the HADS-D showed greater convergence with each other than with the HADS-A. The established severity cut-off scores for the HADS-D and the PHQ-9 were assessed for convergence using Wilcoxon signed-rank test for related samples. Responsiveness to clinical change, from baseline to end of treatment, was measured by running paired *t*-tests on the HADS-D and the PHQ-9. Effect size of both measures was then calculated.

Analyses were carried out using Statistical Package for the Social Sciences (SPSS version 14) and Clinimetrics Toolkit (CMT).

### RESULTS

A total of 1496 patients were referred to the mental health workers' service between February 2005 and March 2006. Subsequently, 1087 (73%) attended the service and were assessed using the PHQ-9 at the initial appointment; 1063 completed baseline HADS before or on the day of first attending; 478 (45%) patients were assessed with the PHQ-9 within 3 days of completing the HADS. To ensure scale responses referred to the same time reference, this smaller sample was used for assessing relative convergent and discriminant validity, convergence of severity banding, and responsiveness to change.

**Table 2. Distribution of participants across the HADS-D and PHQ-9 severity ratings in baseline and end of treatment samples.**

HADS-D	n (%)	PHQ-9	n (%)
Baseline			
None (<8)	151 (33)	None (<5)	36 (8)
Mild (8–10)	139 (30)	Mild (5–9)	81 (18)
Moderate (11–15)	127 (27)	Moderate (10–14)	140 (30)
Severe (>15)	45 (10)	Moderately severe (15–19)	118 (25)
		Severe (>19)	87 (19)
Total	462		462
End of treatment			
None (<8)	390 (79)	None (<5)	246 (50)
Mild (8–10)	66 (13)	Mild (5–9)	124 (25)
Moderate (11–15)	27 (6)	Moderate (10–14)	60 (12)
Severe (>15)	8 (2)	Moderately severe (15–19)	36 (8)
		Severe (>19)	25 (5)
Total	491		491

HADS-D = depression subscale of the Hospital Anxiety and Depression Scale.  
PHQ-9 = Patient Health Questionnaire.

At the end of treatment, 544 patients (50%) completed the HADS and PHQ-9.

### Sample characteristics

Table 1 shows demographic characteristics of service attenders. Most participants were female, employed, and educated beyond minimum school age.

### Reliability

Cronbach's  $\alpha$  coefficients (plus 95% confidence intervals [CIs]) and item total correlations for the HADS-D and PHQ-9 at baseline and end of treatment are shown in Appendix 1. Coefficient  $\alpha$  for both scales are acceptable and comparable across the time points (range = 0.83 to 0.92). As all-item total correlations within the HADS-D and PHQ-9 exceed 0.4, these can all be considered adequate.

### Factor structure

For HADS-D scores at baseline, the first principal component explained 52% of the variance; for the PHQ-9 the corresponding figure was 42%. The item loadings for both scales are shown in Appendix 2. Most items within each scale had a substantial loading, indicating that the HADS-D and the PHQ-9 are both factorially valid. The coefficient of congruence was 0.999 for both the HADS-D and PHQ-9, indicating that the factor structure of both measures is highly robust across time (from baseline to end of treatment).

### Convergent and discriminant validity

Intercorrelations of the HADS-A, the HADS-D, and PHQ-9 at baseline and at end of treatment are shown in Appendix 3. Correlations were all significant at the 0.01 level, as would be expected between such closely-related constructs as anxiety and depression. Using William's test, the correlations were significantly higher between the PHQ-9 and HADS-D than between either of these measures and the HADS-A. Correlations at baseline were: HADS-D with PHQ-9 (0.68) versus HADS-D with HADS-A (0.49),  $P < 0.001$ ; HADS-D with PHQ-9 (0.68) versus PHQ-9 with HADS-A (0.48),  $P < 0.001$ . The same pattern of results was obtained at the end of treatment: HADS-D with PHQ-9 (0.81) versus HADS-D with HADS-A (0.74),  $P < 0.001$ ; HADS-D with PHQ-9 versus PHQ-9 with HADS-A (0.77),  $P = 0.01$ .

### Convergence of severity banding

Table 2 shows the distribution of scores falling within PHQ-9 and HADS-D severity cut-offs. Although both measures purport to measure severity of depressive symptoms, there is a lack of concurrence of distribution within cut-off bands. These differences are significant at baseline ( $P < 0.001$ ) and at end of treatment ( $P < 0.001$ ), indicating that PHQ-9 categorises greater severity of symptoms than HADS-D.

### Responsiveness to change

Paired *t*-tests from baseline to end of treatment indicated a significant change in both the PHQ-9 and the HADS-D, reflecting a reduction in depressive symptoms. The mean score on the PHQ-9 was 12.7 (standard deviation [SD] = 6.47) at baseline, and 6.25 (SD = 6.01) at the end of treatment (95% CI = 5.79 to 7.03). The mean HADS-D score was 8.85 (SD = 4.52) at baseline, and 4.31 (SD = 4.02) at the end of treatment (95% CI = 4.11 to 4.97). The effect size for change on the PHQ-9 was 0.99 compared with 1.0 for the HADS-D, indicating that the scales are comparable in terms of their sensitivity to change.

## DISCUSSION

### Summary of main findings

Both the HADS-D and PHQ-9 demonstrated reliability, convergent/discriminant validity, robustness of factor structure, and responsiveness to change in a sample of primary care patients referred to mental health workers. However, given that both scales purport to measure severity of depression, the level of agreement shown in this regard was disappointing. If treatment decisions are to be made on the basis of severity, this indicates that further work is needed to assess the validity of both scales' endorsed severity cut-off bands.

### Strengths and limitations of the study

This study assessed the psychometric properties of two depression severity rating scales, advocated by the British Medical Association, in a UK sample of primary care patients who GPs had identified as having a mild to moderate mental health problem. Participants had therefore been drawn from the same patient group in which depression severity measurement was intended to apply. This provides useful comparison data to allow practitioners to make a more informed choice than has previously been possible. Moreover, the present analyses are the first to report the factor structure of the PHQ-9 in a UK sample.

Ideally, the study would have included the assessment of the severity bandings of these scales against a clinical 'gold standard' such as the Hamilton Depression Rating Scale,<sup>11</sup> or the Structured Clinical Interview for DSM-IV (SCID);<sup>9</sup> however, that was beyond the scope of the present assessment where data were primarily collected for purposes of an audit.

Part of the inclusion criteria for referral to a mental health worker was the identification of a 'mild to moderate' mental health problem. This required GPs to make their own assessment as to whether a patient fitted this criterion before initiating the referral. The scales were only completed once referral had been made. The fact that both scales categorised some of those patients as having severe depression highlights the difficulty faced when following a clinical

impression of severity alone, which has previously been shown to not always be reliable.<sup>12</sup> However, the disparity between the measures also demonstrates that either one, or both, of these measures is categorising depression severity erroneously.

At baseline, method variance may have explained some differences in severity categorisation between the PHQ-9 and HADS-D, as the PHQ-9 was conducted as an interview at this stage while the HADS-D was administered as a self-complete questionnaire. The difference in methods arose because the mental health workers found it useful to include the PHQ-9 as an interview at the first assessment. This is acceptable for the psychometric assessment because the PHQ-9 has demonstrated concurrent diagnostic validity for self-complete and interview-administered methods.<sup>13</sup> It should therefore be possible to use these methods interchangeably.

Possible method variance may have also occurred due to a time delay of up to 3 days in completion of both scales at baseline. As the time-reference point for the scales would have overlapped, it was considered acceptable to include data collected up to 3 days apart. The difference in severity categorisation remained at the end of treatment (at which time and administration methods were the same), which further refutes the likelihood of method variance explaining the difference.

The proportion of patients completing the questionnaire at the end of treatment was only half that of patients completing at baseline. The reduced number reflects the fact that the way people disengage from services and completion of questionnaires inevitably relies on postal return. Postal reminders and reply-paid envelopes were used. The response rate was comparable with other studies requiring postal return.<sup>14</sup>

### Comparison with existing literature

In keeping with previous investigations of the PHQ-9,<sup>4,15</sup> and HADS-D,<sup>16,17</sup> both scales exhibited good internal consistency. The factor structure of the HADS-D also reflects investigations on the scale in general,<sup>18</sup> and clinical samples where the depression subscale emerges within the overall HADS items.<sup>19</sup> The factor structure of the PHQ-9 has been reported in a US sample, where the variance explained by a single factor, including all the PHQ-9 items, ranged from 39% to 49% across different ethnic groups.<sup>15</sup> This is comparable with the current analyses where 42% of the variance was explained.

The differences found between the HADS-D and the PHQ-9 in the distribution of scores by severity banding is a concern. Lowe *et al* assessed the relative validity of the PHQ-9, HADS-D, and the World Health Organization Well-Being Index against



the SCID, focusing on cut-off points of cases/non-cases of 'major depression' and of 'any depressive disorder'.<sup>20</sup> They did not report on the comparative severity bandings within cases against the SCID results. However, their findings do deviate from the endorsed cut-off scores for cases/non-case: for major depression a cut-off point of  $\geq 11$  was recommended for the PHQ-9 and  $\geq 9$  for the HADS-D. For 'any depressive disorder' the cut-off for the HADS-D remained  $\geq 8$  but was  $\geq 9$  for the PHQ-9, suggesting some over-inclusion in the original cut-offs, particularly with regard to the PHQ-9.

### Implications for future research or clinical practice

Further research is required to investigate the psychometric properties of the PHQ-9, HADS, and BDI-II in a UK sample, including a concurrent validation assessment against an ICD-10 clinical interview; from this, empirically-derived severity cut-offs could be established.

Although NICE guidelines emphasise the importance of considering other factors in addition to severity when looking at treatment options,<sup>1</sup> the fear has been that some policy developments may favour a more prescriptive approach. In Grampian, the vast majority of practices have opted for the PHQ-9 for assessing depression severity. In the present sample, if the stepped-care model were to be applied rigidly,<sup>1</sup> 74% of patients assessed with the PHQ-9 would be offered an antidepressant. However, had the same sample been assessed with the HADS-D, only 37% would fall within the prescribing category. A recent Scottish Executive guide, which advocated use of the PHQ-9, indicated that patients without a history of depression and with a score  $< 15$  should not be prescribed an antidepressant.<sup>21</sup> It is of concern that clinicians are being advised to follow such a rigid code. Presently, clinicians should exercise caution in interpreting scores according to the endorsed severity cut-offs for the HADS-D or PHQ-9.

### Funding body

Data were collected as part of an audit of the 'Doing Well by People with Depression' pilot, funded by the Centre for Change and Innovation, Scottish Executive

### Ethical approval

The Scientific Adviser of Grampian Research Ethics Committee advised that the present analysis of this audit data did not require ethical approval so long as the data were anonymous

### Competing interests

The authors have stated that there are none

### Acknowledgements

The authors would like to thank all the patients who completed questionnaires, the nine Grampian-based mental health workers who undertook the data collection, and the 32 general practices in Grampian that participated in the 'Doing Well by People with Depression' pilot

### Discuss this article

Contribute and read comments about this article on the Discussion Forum: <http://www.rcgp.org.uk/bjgp-discuss>

## REFERENCES

1. National Institute for Clinical Excellence. *Depression: management of depression in primary and secondary care (NICE guideline). Clinical guideline 23*. London: National Institute for Clinical Excellence, 2004.
2. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*. 10th revision. Geneva: World Health Organization, 2006. <http://www.who.int/classifications/apps/icd/icd10online/> (accessed 9 Aug 2007).
3. British Medical Association. *Quality and outcomes framework*. London: British Medical Association, 2006. <http://www.bma.org/ap.nsf> (accessed 9 Aug 2007).
4. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**(9): 606–613.
5. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**(6): 361–370.
6. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *J Personal* 1996; **67**(3): 588–597.
7. Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004; **78**(2): 131–140.
8. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract* 2007; **57**(535): 144–151.
9. First MB, Spitzer RL, Williams JBW, Gibbon M. *Structured Clinical Interview for DSM-IV (SCID)*. Washington DC: American Psychiatric Press Inc., 1995.
10. Wrigley C, Neuhaus JO. The matching of two sets of factors. *Am Psychol* 1955; **10**: 418–419.
11. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; **23**: 56–62.
12. Kendrick T, King F, Albertella L, Smith PWF. GP treatment decisions for patients with depression. *Br J Gen Pract* 2005; **55**(513): 280–286.
13. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; **282**(18): 1737–1744.
14. Cameron IM, Cunningham L, Crawford JR, et al. Psychometric properties of the BASIS-24<sup>®</sup> (Behaviour and Symptom Identification Scale-Revised) Mental Health Outcome Measure. *Int J Psychiatr Clin Pract* 2007; **11**: 36–43.
15. Huang FY, Chung H, Kroenke K, et al. Using the Patient Health Questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *J Gen Intern Med* 2006; **21**(6): 547–552.
16. Bjelland I, Dahl AA, Tangen Haug T, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *J Psychosom Res* 2002; **52**(2): 69–77.
17. Crawford JR, Henry JD, Crombie C, Taylor EP. Normative data for the HADS from a large non-clinical sample. *Br J Clin Psychol* 2001; **40**(Pt 4): 429–434.
18. Dunbar M, Ford G, Hunt K, Der J. A confirmatory factor analysis of the Hospital Anxiety and Depression Scale: comparing empirically and theoretically derived structures. *Br J Clin Psychol* 2000; **39**(Pt 1): 79–94.
19. Dagnan D, Chadwick P, Trower P. Psychometric properties of the Hospital Anxiety and Depression Scale with a population of members of a depression self-help group. *Br J Med Psychol* 2000; **73**(Pt 1): 129–137.
20. Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004; **78**: 131–140.
21. Scottish Executive. *Delivering for health: delivering for mental health. Using self-help in primary care and community based services. A guide to everyday service delivery for mild to moderate psychological problems*. Edinburgh: Scottish Executive, 2006. <http://www.scotland.gov.uk/Publications/2007/01/09114002/0> (accessed 13 Aug 2007).

### Appendix 1. Cronbach's $\alpha$ and item-total correlations of the HADS-D and PHQ-9 in baseline and end of treatment samples.

Scale and items	Baseline			End of treatment		
	<i>n</i>	Coefficient $\alpha$ (95% CI)	Item total correlations	<i>n</i>	Coefficient $\alpha$ (95% CI)	Item total correlations
HADS-D	1030	0.84 (0.83, 0.85)		523	0.89 (0.87 to 0.9)	
Still enjoy			0.67			0.77
Laugh			0.67			0.73
Cheerful			0.65			0.74
Slowed down			0.51			0.54
Appearance			0.47			0.58
Look forward			0.69			0.81
Enjoy book/radio/TV			0.55			0.63
PHQ-9	1078	0.83 (0.81, 0.84)		507	0.92 (0.91 to 0.93)	
Little interest			0.59			0.75
Feeling down			0.65			0.79
Trouble sleeping			0.50			0.69
Tired			0.54			0.70
Poor appetite			0.49			0.68
Feeling bad about self			0.54			0.78
Trouble concentrating			0.56			0.74
Slowing			0.50			0.66
Thoughts of self-harm			0.42			0.62

HADS-D = depression subscale of the Hospital Anxiety and Depression Scale. PHQ-9 = Patient Health Questionnaire.

### Appendix 2. Factor analysis item loadings on the HADS-D and PHQ-9.

HADS-D item loadings ( <i>n</i> = 1030)		PHQ-9 item loadings ( <i>n</i> = 1078)	
Item	Loading	Item	Loading
Still enjoy	0.79	Little interest	0.71
Laugh	0.79	Feeling down	0.77
Cheerful	0.77	Trouble sleeping	0.61
Slowed down	0.63	Tired	0.65
Appearance	0.59	Poor appetite	0.61
Look forward	0.81	Feeling bad about self	0.67
Enjoy book/ radio/TV	0.67	Trouble concentrating	0.68
		Slowing	0.61
		Thoughts of self-harm	0.54

HADS-D = depression subscale of the Hospital Anxiety and Depression Scale. PHQ-9 = Patient Health Questionnaire.

### Appendix 3. Intercorrelations of the HADS-A, HADS-D, and PHQ-9.<sup>a</sup>

Scale	HADS-A	HADS-D	PHQ-9
At baseline ( <i>n</i> = 454)			
HADS-A	–	0.49	0.48
HADS-D	–	–	0.68
PHQ-9	–	–	–
At end of treatment ( <i>n</i> = 491)			
HADS-A	–	0.74	0.77
HADS-D	–	–	0.81
PHQ-9	–	–	–

<sup>a</sup>*P* < 0.01 for all correlations. HADS-A = anxiety subscale of the Hospital Anxiety and Depression Scale. HADS-D = depression subscale of the HADS. PHQ-9 = Patient Health Questionnaire